



基于交互序列商品相关性建模的图卷积会话推荐

闫昭, 项欣光, 李泽超*

南京理工大学计算机科学与工程学院, 南京 210094

* 通信作者. E-mail: zechao.li@njust.edu.cn

收稿日期: 2021-03-04; 修回日期: 2021-04-01; 接受日期: 2021-06-03; 网络出版日期: 2022-06-10

科技创新 2030 “新一代人工智能”重大项目课题 (批准号: 2018AAA0102002)、国家自然科学基金企业创新发展联合基金重点支持项目 (批准号: U20B2064) 和 “万人计划” 青年拔尖人才项目资助

摘要 会话推荐旨在根据用户在会话中的短期行为, 预测用户的后续行为, 多应用于用户匿名访问网站的场景. 会话推荐的核心在于商品之间协同关系的建模, 以提升推荐结果的性能. 现有方法大都挖掘商品交互顺序信息以计算商品间的协同关系. 这种方式仅仅考虑了商品之间的相邻交互的信息, 忽视了历史交互信息. 其实, 若两个商品共享相同的历史交互信息, 即它们具有一定的协同关联性. 例如, 在两个商品交互序列为 $I_1-I_2-I_3$ 和 $I_1-I_2-I_4$ 的会话中, 商品 I_3 与 I_4 的历史交互信息均为 I_1-I_2 , 则商品 I_3 与 I_4 极有可能具有较强的关联关系. 因此, 本文提出了一种新颖的会话推荐模型 IHGCN, 通过分析会话的商品交互序列建模商品之间的关联关系, 同时挖掘了相邻交互信息和历史交互信息. 为了从这些交互信息中协同分析出商品特征, 所提出方法使用图结构构建商品之间的关系, 并引入图卷积网络从商品关系图中学习商品特征. 该方法使用特征维度级的细粒度注意力机制挖掘用户的全局偏好, 并融合用户的局部偏好; 基于学习得到的商品特征和用户偏好特征生成最终的推荐结果. 本文在 3 个公开数据集上进行了验证分析, 与当前最好的相关方法的比较结果表明所提出方法 IHGCN 的有效性和优越性.

关键词 会话推荐, 图卷积网络, 商品交互相似性, 用户偏好, 注意力机制

1 引言

各类社交网络平台存在着大量的匿名用户. 在这些匿名用户访问时, 社交平台为他们创建一个临时会话窗口提供服务. 当用户离开后, 该会话窗口随之关闭, 并会永久失效. 社交网络平台只能在匿名用户的短暂操作过程中获取到用户信息, 并且需要根据这些短暂交互行为快速确定用户偏好, 为用户更好地提供服务. 这种推荐任务被称之为会话推荐. 由于会话周期短, 同时缺乏用户的全局交互信息, 经典的推荐算法思想^[1~4]无法有效地从用户的短期历史行为中分析出用户偏好, 因此无法直接用于

引用格式: 闫昭, 项欣光, 李泽超. 基于交互序列商品相关性建模的图卷积会话推荐. 中国科学: 信息科学, 2022, 52: 1069–1082, doi: 10.1360/SSI-2020-0383

Yan Z, Xiang X G, Li Z C. Item correlation modeling in interaction sequence for graph convolutional session recommendation (in Chinese). Sci Sin Inform, 2022, 52: 1069–1082, doi: 10.1360/SSI-2020-0383

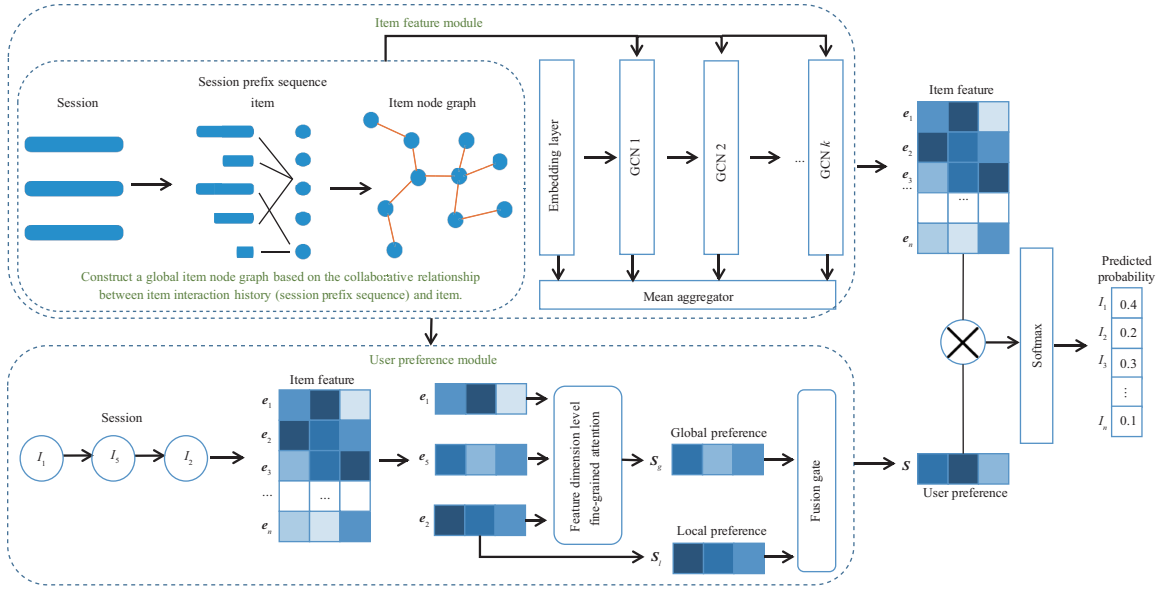


图 1 (网络版彩图) IHGCN 模型框架图. IHGCN 包含商品特征模块和用户偏好模块. 商品特征模块基于图卷积网络学习商品特征. 用户偏好模块通过融合用户的全局偏好和局部偏好获取用户的偏好特征

Figure 1 (Color online) Framework of the proposed IHGCN. IHGCN contains the item feature module and user preference module. The item feature module learns feature representations of items based on GCN, while the user preference module uncovers the user preference by integrating the global and local preference of users

解决会话推荐任务. 于是, 会话推荐作为推荐系统^[5,6]中的一类新任务, 近年来颇受学术界和工业家关注.

在早期的会话推荐工作中, Item-KNN^[7]使用协同过滤方法^[8~11]解决会话推荐问题. Item-KNN 忽视了会话中其他商品对用户偏好的贡献, 难以准确捕获用户偏好. 基于马尔可夫 (Markov) 链的会话推荐模型^[12,13]假设下一个状态的概率分布仅取决于当前状态, 同样忽视了会话中历史商品对用户偏好的贡献. 近年来, 深度学习模型在会话推荐中逐渐发挥出巨大的能量. GRU4REC^[14]首次在会话推荐中使用递归神经网络 (recurrent neural network, RNN)^[15,16], 取得了较好的效果. 在此基础上, 多层感知机 (multilayer perceptron, MLP) 和粗粒度注意力机制也被引入到会话推荐中^[17,18]. 文献^[19,20]使用了分级 RNN. RNN-KNN^[21]在获取当前会话的用户偏好时增加了邻居会话的用户偏好. 这些方法大都使用粗粒度注意力机制融合特征向量, 忽视了特征向量分量的贡献差异性.

图神经网络^[22]实现了图结构中信息的传播, 表现出不凡的效果. NGCF^[23]和 LightGCN^[3]将图神经网络引入到推荐任务中. NGCF 根据用户和商品间的协同关系构建图, 使用 GCN 提取用户特征和商品特征. LightGCN 是 NGCF 的轻量化版本, 改进了生成用户特征和商品特征的方法, 简化了训练过程. 在会话推荐中, SR-GNN^[24]基于图神经网络为每一个商品交互序列构建局部图结构. 但该方法构建的图只是商品节点整体图的子图, 忽视了商品节点间固有的关联性.

鉴于此, 本文提出了一种新颖的图卷积会话推荐模型, 并将其命名为 IHGCN (interactive and historical graph convolutional network). 所提出方法的框架图如图 1 所示. 该模型挖掘商品在交互序列中体现的相关性构建图结构, 利用图卷积网络学习项目的特征表示, 进一步整合历史会话序列得到会话表示, 最终达成推荐目标. 具体的来讲, 所提出模型 IHGCN 包含商品特征模块和用户偏好模块. 商品特征模块挖掘商品的历史交互信息学习得到商品的特征表示. 如果两个商品具有相同的历史交互信

息,则认为它们有着一定的协同相关性.例如,在两个商品交互序列为 $I_1-I_2-I_3$ 和 $I_1-I_2-I_4$ 的会话中,商品 I_3 与 I_4 的历史交互信息均为 I_1-I_2 ,则认为,商品 I_3 与 I_4 极有可能具有较强的关联.为了更好地抽取出商品特征,商品特征模块采用图结构建模所有这些关联关系,并利用图卷积网络从中学习得到商品特征.此外,IHCG使用了均值增加器融合商品隐层特征和图卷积网络的各层输出特征.为了更好地建模用户偏好,用户偏好模块融合了用户的全局偏好特征和局部偏好特征.为了挖掘用户的全局偏好,用户偏好模块引入了特征维度级的细粒度注意力机制,差异化融合特征向量的分量,从会话序列中的全部商品信息中学习得到用户的全局偏好特征.用户的局部偏好特征由当前会话中的最新交互决定.最后,商品推荐结果根据得到的商品特征和用户偏好特征采用 Softmax 操作获得.为了验证所提出方法在会话推荐任务中的性能,本文在 3 个公开的数据集上进行了实验比较分析,即 Yoochoose 数据集、Diginetica 数据集和 Globo 数据集.与最新的方法相比,本文所提出方法 IHGCN 取得了优越的性能,表明了其有效性.

本文的主要贡献总结如下:

(1) 本文提出一种新颖的会话推荐算法 IHGCN,综合考虑了商品的相邻交互信息和历史交互信息,以及用户的全局偏好和局部偏好.这是会话推荐任务中第 1 次显式挖掘商品的历史交互信息的工作.

(2) 本文提出由商品交互历史与商品之间的协同关系构建商品节点图,能够更准确地捕获商品之间的关联关系.

(3) 本文提出通过特征维度级的细粒度注意力机制融合会话中的商品特征,解析特征不同维度信息的差异化,并融合最新交互信息得到准确的用户偏好特征表示.

本文的组织结构如下.第 2 节阐述相关工作,包含了会话推荐和图神经网络两部分.在第 3 节中,我们将详细介绍本文所提出方法.相关实验结果和分析将在第 4 节中给出.最后,第 5 节对本文进行了总结.

2 相关工作

本节回顾一些会话推荐和图神经网络的相关工作.相关工作分为两个部分,分别为会话推荐和图神经网络.

2.1 会话推荐

Item-KNN^[7]和矩阵分解^[25~27]方法常用来解决会话推荐任务. Item-KNN 方法首先构建会话与商品的交互矩阵,使用交互矩阵商品对应向量间的余弦值度量商品间的相似性,最后推荐与会话最新交互商品相似的商品.矩阵分解模型把会话与商品的交互矩阵分解为会话的潜在因子矩阵和商品的潜在因子矩阵,使用商品交互序列中所有商品潜在因子向量的均值估计用户偏好.随后一些工作探索使用马尔可夫链解决会话推荐问题^[12,13].马尔可夫链的假设是下一个状态的概率分布仅取决于当前状态.这些方法忽视了会话中历史商品对用户偏好的贡献,很难精准地挖掘用户偏好.

随着深度学习^[28]的崛起,不少工作提出使用深度学习思想构建会话推荐模型.这些方法大都通过学习模型挖掘用户偏好,最终使用交叉熵函数预测推荐结果.根据是否融合用户局部偏好,相关方法可分为两类.一类方法仅挖掘用户全局偏好,不融合用户局部偏好,如 GRU4REC^[14], Improved-RNN^[29]和 RNN-KNN^[21]方法. Improved-RNN^[29]意图在 RNN 的基础上进行数据强化.此类方法把商品交互序列依次输入到 RNN 模型中,得到的隐藏状态即为用户偏好特征.另一类方法挖掘用户局部偏好,

并把全局偏好与局部偏好的融合向量作为用户偏好特征^[30]. 在生成全局偏好的过程中, 常用的方法是使用 RNN 处理商品交互序列生成隐藏状态, 如 NARM^[17]. 另一类做法使用粗粒度注意力机制融合商品交互序列的所有商品, 如 STAMP^[18] 和 SR-GNN^[24]. SR-GNN 引入了门控图神经网络 (gated graph neural networks, GGNN)^[31], 使用 GGNN 生成商品特征. 在生成局部偏好的过程中, 常用的方法是联合使用 RNN 和粗粒度注意力机制, 使用经过 RNN 处理的部分商品交互序列生成的隐藏状态的注意力叠加向量, 如 NARM^[17]. 另一类方法是使用会话的最后一项, 如 STAMP^[18] 和 SR-GNN^[24]. 虽然这些方法在会话推荐中实现了良好的效果, 但是它们一方面忽视了商品节点的全局邻接关系, 另一方面忽视了特征向量分量的贡献差异性.

为此, 本文首次显示挖掘商品的历史交互信息, 根据商品及其对应的历史交互信息, 提出了分析商品节点的全局邻接关系的会话推荐方法, 并使用特征维度级的细粒度注意力机制解析特征矢量不同维度的贡献差异性.

2.2 图神经网络

近些年, 不少工作研究如何提取图结构数据的节点特征, 它们所使用的处理图结构数据的这一类神经网络称之为图神经网络. 这里介绍几种常见的图卷积神经网络. GCN^[32] 首先根据节点的邻接关系构建邻接矩阵, 使用邻接矩阵与全部节点特征向量组成的矩阵乘积来叠加节点的邻居特征, 叠加之后经过非线性变换得到输出. GAT^[33] 的想法与 GCN 类似, 使用注意力机制自适应学习邻居节点间的权重. 不同于 GCN, GAT 节点的邻居权重是通过注意力机制得到的, 权重随节点特征的变化而变化, 而 GCN 节点的邻居权重为固定值. GraphSAGE^[34] 同样是叠加节点的邻居特征. 为了避免全图训练, 它对节点的邻居进行了取样操作, 使每个节点都有固定数量的邻居, 从而可以小批次训练. NGCF^[23] 把 GCN 引入了推荐系统, 它把用户与商品的交互矩阵作为邻接矩阵, 并使用 GCN 生成用户特征和商品特征, 用户特征为他交互过的商品特征的叠加, 商品特征为它交互过的用户特征的叠加. LightGCN^[3] 在 NGCF 上进行改进, 保留了 GCN 叠加节点的邻居特征这一想法, 舍弃了 GCN 的非线性变换. 图卷积神经网络善于提取图结构数据的节点特征, 因而本文引入图卷积神经网络分析会话的历史交互信息, 以提取更好的商品节点特征.

3 方法

本节将详细介绍本文所提出的会话推荐方法, 包含问题描述、模型概览, 以及具体实现细节等.

3.1 问题描述

会话推荐涉及到商品、会话等. 在本文中, $\mathcal{I} = \{I_1, I_2, \dots, I_N\}$ 表示包含 N 个商品的集合. 假设存在一个长度为 L 的商品交互序列 $s = [I_{s,1}, I_{s,2}, \dots, I_{s,L}]$, $I_{s,j} \in \mathcal{I}$, 会话推荐根据用户在 s 中交互过的商品预测用户偏好, 进而计算所有待预测商品的点击概率 $\hat{p}$, 最后推荐概率最大的 k 个商品. 本文涉及的重要符号定义详见表 1 所示.

3.2 模型概览

本文所提出 IHGCN 方法如图 1 所示. IHGCN 包含了商品特征模块和用户偏好模块. 在商品特征模块, IHGCN 首先根据商品交互历史与商品间的协同关系构建全局的商品节点图, 并引入图卷积神经网络从商品节点图中学习得到商品特征表示. 为了更好地获取商品的信息, 均值增加器被引入将商品的隐特征和各图卷积网络层输出进行融合, 生成商品的特征表示. 用户偏好模块联合分析了用户的全局

表 1 符号定义

Table 1 Definition of notation

Notation	Description	Notation	Description	Notation	Description
$\mathcal{I}$	Item set	$\mathcal{H}$	Session prefix set	I	Item
$\mathbf{A}$	Adjacent matrix of items	$\mathbf{D}$	Diagonal matrix	α_k	Weight
$\mathbf{v}$	Adjacent vector of items	$\mathbf{e}$	Item feature vector	$\mathbf{s}_g$	Global preference
$\mathbf{s}_u$	User preference	$\mathbf{W}$	Parameter matrix	$\mathbf{b}$	Parameter vector
$\hat{p}$	Prediction probability	$\mathcal{L}$	Loss function	$\mathcal{G}$	Item node graph
$\mathbf{E}^{(k)}$	Output of graph convolution	$\mathbf{s}_l$	Local preference	$\hat{z}$	Item score
s	Item interaction sequence				

偏好和局部偏好信息以更好地建模用户偏好. 为了挖掘用户的全局偏好, 特征维度级的细粒度注意力机制被用来从与用户交互的全部商品特征中学习用户的偏好. 而用户局部偏好是从会话最新交互项的商品特征中获取得到. 基于商品特征和用户偏好特征, IHGCN 引入了 softmax 层生成推荐结果.

3.3 商品节点图

3.3.1 会话前缀序列与商品之间的协同矩阵

如图 2(a) 所示, 在会话推荐中, 若两个商品具有相同的历史交互信息, 即可认为它们有着一定的协同相似性, 因而商品之间的协同关系可由它们对应的商品交互历史来确定. 商品交互历史为上一时刻的会话前缀序列, 例如: 会话 $s_2 = [I_1, I_2, I_3]$, $I_j \in \mathcal{I}$ 是长度为 3 的商品交互序列, 商品 I_3 的会话前缀序列 $[I_1, I_2]$. 统计所有商品的会话前缀序列, 构成了包含 M 个会话前缀序列的会话前缀集 $\mathcal{H} = \{h_1, h_2, \dots, h_M\}$. 根据会话前缀集和商品集, 可以确定会话前缀序列与商品的协同矩阵, 其中矩阵分量的值表示相应会话前缀序列与商品共同出现的次数.

3.3.2 商品邻接矩阵

如图 2(b) 所示, 根据上述协同矩阵中商品的对应行, 可以确定每一个商品向量 $\mathbf{v}_{I_i} \in \mathbb{Z}^{|\mathcal{H}|}$, 其中 $|\mathcal{H}|$ 表示会话前缀集的大小. 最终通过余弦相似度计算商品之间的协同相似性. 依据上述关系构建一个商品节点图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, 图中的节点为商品集, 图中的边为商品之间的相似度. 该图使用邻接矩阵 $\mathbf{A}$, $\mathbf{A} \subseteq \mathbb{R}^{N \times N}$ 进行存储表示, 其中 $\mathbf{A}_{ij}$ 表示商品 I_i 和 I_j 之间的相似度, 计算公式如下:

$$\mathbf{A}_{ij} = \frac{\mathbf{v}_{I_i} \cdot \mathbf{v}_{I_j}}{\|\mathbf{v}_{I_i}\| \|\mathbf{v}_{I_j}\|}, \quad (1)$$

其中 $\cdot$ 表示向量内积运算, $\|\cdot\|$ 表示向量的模. $\mathbf{v}_{I_i} \in \mathbb{R}^{|\mathcal{S}|}$ 表示商品 I_i 对应的向量, 向量维度为会话集的大小, 向量分量为对应商品交互历史下一次点击为 I_i 的次数. 另外, $\mathbf{A}_{ii} = 1$ 表示商品 I_i 自身的相似度.

3.4 商品特征

构建商品节点图 $\mathcal{G}$ 后, 本小节介绍如何通过图卷积网络 (GCN) 生成商品特征. GCN 的一般形式如下:

$$\mathbf{E}^{(k+1)} = \text{gcn}(\mathbf{E}^{(k)}), \quad (2)$$

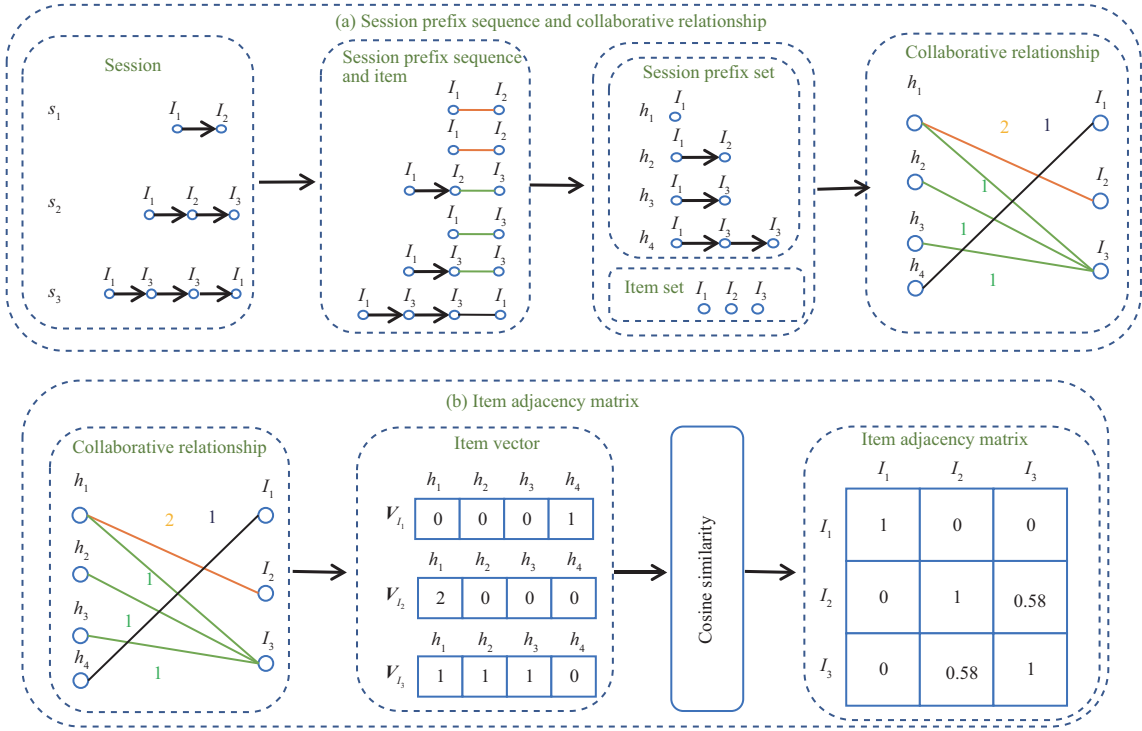


图 2 (网络版彩图) 商品节点相似度

Figure 2 (Color online) Illustration of item node similarity. (a) Session prefix sequence and collaborative relationship; (b) item adjacency matrix

其中, $\mathbf{E}^{(k)} \in \mathbb{R}^{N \times D}$ 表示第 k 层图卷积的输出, D 表示输出特征维度, $\mathbf{E}^{(0)}$ 表示第 1 层图卷积的输入, 由各商品隐特征向量构成的矩阵, 是生成商品特征模块的全部训练参数. 随着图卷积层数的增加, 网络可以叠加高阶邻居商品节点的特征, 这有利于提取更为准确的商品特征. 本文 GCN 的具体形式如下:

$$\text{gcn}(\mathbf{E}^{(k)}) = \mathbf{D}^{-1} \mathbf{A} \mathbf{E}^{(k)}, \quad (3)$$

其中 $\mathbf{A} \in \mathbb{R}^{N \times N}$ 是商品的邻接矩阵. $\mathbf{D} \in \mathbb{R}^{N \times N}$ 是一个对角矩阵, $D_{ii} = \sum_{j=1}^N A_{ij}$. $\mathbf{D}^{-1} \mathbf{A}$ 是拉普拉斯 (Laplace) 矩阵的一种形式, 对邻接矩阵 $\mathbf{A}$ 进行行归一化操作. 生成商品特征的公式如下:

$$\mathbf{E} = \sum_{i=0}^k \alpha_i \mathbf{E}^{(i)}, \quad (4)$$

其中, $\mathbf{E} \in \mathbb{R}^{N \times D}$ 表示商品特征. $\alpha_i > 0$ 为第 i 层图卷积输出对商品特征的重要性, 本文设置 $\alpha_i = 1/(k+1)$.

3.5 用户偏好

本文的用户偏好 s_u 自适应的融合商品交互序列的全局偏好 s_g 和局部偏好 s_l . 对任意一个会话 $s = [I_{s,1}, I_{s,2}, \dots, I_{s,L}]$, $s \in \mathcal{S}$, $I_{s,j} \in \mathcal{I}$. s 对应的商品特征序列为 $[e_1, e_2, \dots, e_L]$, $e_i \in \mathbb{R}^D$ 是经过图卷积网络生成的商品特征, 用户偏好的计算公式如下:

$$s_u = \eta \odot s_g + (1 - \eta) \odot s_l. \quad (5)$$

式中 $\boldsymbol{\eta} \in \mathbb{R}^D$ 控制着全局偏好和局部偏好的重要性, 具体计算如下:

$$\boldsymbol{\eta} = \sigma(\mathbf{W}_3 \mathbf{s}_g + \mathbf{W}_4 \mathbf{s}_l + \mathbf{b}_u), \quad (6)$$

其中, $\mathbf{W}_3 \in \mathbb{R}^{D \times D}$, $\mathbf{W}_4 \in \mathbb{R}^{D \times D}$, $\mathbf{b}_u \in \mathbb{R}^D$ 表示可训练参数. D 表示商品特征维度. $\odot$ 为向量按位相乘的结果. σ 为非线性激活函数. 下面详细介绍如何计算用户全局偏好 $\mathbf{s}_g$ 和局部偏好 $\mathbf{s}_l$.

全局偏好. 全局偏好 $\mathbf{s}_g$ 叠加了商品交互序列的全部商品特征. 公式如下:

$$\mathbf{s}_g = \sum_{i=1}^L \gamma_i \odot \mathbf{e}_i, \quad (7)$$

其中 $\gamma_i \in \mathbb{R}^D$ 表示商品交互序列第 i 个位置商品的重要性. 为了差异化融合商品特征分量, 本文使用特征维度级的细粒度注意力机制. 另外, 商品位置是影响商品重要性的一个因素, 因而生成全局偏好的过程考虑了商品位置的影响. 公式如下:

$$\boldsymbol{\beta}_i = \sigma(\mathbf{W}_{i,1} \mathbf{e}_i + \mathbf{W}_{i,2} \mathbf{e}_L + \mathbf{b}_i), \quad (8)$$

$$\gamma_i = \frac{\exp(\boldsymbol{\beta}_i)}{\sum_{j=1}^L \exp(\boldsymbol{\beta}_j)}, \quad (9)$$

其中, $\boldsymbol{\beta}_i \in \mathbb{R}^D$. $\exp$ 表示指数函数. $\mathbf{W}_{i,1} \in \mathbb{R}^{D \times D}$, $\mathbf{W}_{i,2} \in \mathbb{R}^{D \times D}$, $\mathbf{b}_i \in \mathbb{R}^D$ 为商品交互序列第 i 个位置的可训练参数. $\mathbf{e}_i \in \mathbb{R}^D$, $\mathbf{e}_L \in \mathbb{R}^D$ 分别为商品交互序列第 i 个位置的商品特征和最新交互项的商品特征. D 表示商品特征维度. 除法为向量按位相除的结果. $\odot$ 为向量按位相乘的结果. σ 为非线性激活函数.

局部偏好. 本文的局部偏好 $\mathbf{s}_l$ 为商品交互序列最新交互项的商品特征. 即

$$\mathbf{s}_l = \mathbf{e}_L. \quad (10)$$

3.6 评价得分与模型训练

生成用户偏好后, IHGCN 使用用户偏好 $\mathbf{s}_u \in \mathbb{R}^D$ 和待预测商品 $\mathbf{e}_i \in \mathbb{R}^D$ 的内积获得评分, 公式如下:

$$\hat{z}_i = \mathbf{s}_u^T \mathbf{e}_i. \quad (11)$$

之后, 本文使用 softmax 函数获得模型的输出 $\hat{\mathbf{p}} \in \mathbb{R}^N$,

$$\hat{\mathbf{p}} = \text{softmax}(\hat{\mathbf{z}}), \quad (12)$$

其中, $\hat{\mathbf{z}} \in \mathbb{R}^N$ 表示所有待预测商品的评分. $\hat{\mathbf{p}} \in \mathbb{R}^N$ 表示所有待预测商品成为下一次点击的概率. 本文使用交叉熵损失函数, 公式如下:

$$\mathcal{L}(\hat{\mathbf{p}}) = - \sum_{i=1}^N p_i \log(\hat{p}_i) + (1 - p_i) \log(1 - \hat{p}_i), \quad (13)$$

其中, $\mathbf{p} \in \mathbb{R}^N$ 表示商品交互序列下一次点击商品的真实概率分布, 实际上是一位有效编码向量, 即商品交互序列下一次点击的商品概率为 1, 其他商品概率为 0. p_i 为待预测商品对应的概率值, 此值非 0 即 1. 最后, 本文使用 Adam^[35] 优化器优化模型的所有可训练参数.

表 2 数据集统计信息
Table 2 Statistics of the used datasets

	Yoochoose	Diginetica	Globo
Number of clicks	557248	982961	2111251
Number of training sessions	369859	719470	1316937
Number of test sessions	55898	60858	50422
Number of items	16766	43097	8617

4 实验

本节首先介绍文中使用的数据集、基准方法和评价指标, 然后设置了一系列实验, 实验尝试解答以下问题:

- 与基准方法相比, IHGCN 表现如何?
- IHGCN 中不同图卷积层数的表现如何?
- 不同注意力机制的实验效果对比?
- 会话中不同长度序列对推荐结果的影响?

4.1 数据集

本文使用 3 个公开的会话推荐数据集: Yoochoose 数据集¹⁾、Diginetica 数据集²⁾和 Globo 数据集³⁾. 数据集的预处理方法参考 SR-GNN [24]. Yoochoose 是 2015 年 RecSys 挑战赛的数据集, 本文只使用点击文件. Diginetica 是 2016 年 CIKM 挑战赛的数据集, 本文只使用交易文件. Globo 是巴西新闻门户网站数据集, 本文只使用点击文件. 这 3 个数据集都可以在网站上直接下载. 本文按照用户在会话中点击商品的顺序把会话切分为多个商品交互序列, 序列的标签是其对应的下一个商品. 本文选取验证集的方法与 SR-GNN [24] 相同. 数据集的统计结果见表 2.

Yoochoose. 本文对该数据集进行了预处理, 过滤掉长度为 1 的会话和出现次数小于 5 的商品. 过滤后, 数据集包含 7981580 个会话和 37483 个商品. 测试集为最后一天的会话. 训练集为离测试集时间最近的 1/64 时间切片.

Diginetica. 所使用的数据集将原始数据长度为 1 的会话和出现次数小于 5 的商品进行了过滤. 过滤后, 该数据集包含 204771 个会话和 43097 个商品. 最后 7 天的会话作为测试集, 而训练集为剩余的其他会话.

Globo. 该数据集包含前 16 天的交互数据信息. 数据过滤掉长度为 1 的会话和出现次数小于 5 的商品. 过滤后, 数据集包含 743892 个会话和 8617 个商品. 测试集为最后 5 天的会话, 训练集为剩余的其他会话.

4.2 比较方法

为了更好地说明本文所提出方法的有效性, 实验中将本文所提出方法与以下方法进行了比较分析.

- **POP.** 根据商品在数据集中的出现频次进行推荐的方法.

1) <https://www.kaggle.com/chadgostopp/recsys-challenge-2015>.

2) <https://competitions.codalab.org/competitions/11161>.

3) <https://www.kaggle.com/gspmoreira/news-portal-user-interactions-by-globocom/data?select=clicks>.

- **S-POP**. 该方法根据商品在商品交互序列中出现的频次进行推荐.
- **Item-KNN**^[7]. 该方法把商品交互序列的最新交互商品视为用户偏好, 并推荐与之相似的商品.
- **BPR-MF**^[1]. 该方法使用随机梯度下降算法优化成对目标函数.
- **GRU4REC**^[14]. 该方法使用 RNN 处理商品交互序列生成的隐藏状态作为用户偏好, 待预测商品评分是用户偏好经全连接层计算的结果.
- **NARM**^[17]. 该方法使用商品交互序列全局偏好和局部偏好的叠加向量作为用户偏好, 全局偏好是 RNN 处理商品交互序列生成的隐藏状态, 局部偏好是 RNN 处理部分商品交互序列生成的隐藏状态的注意力叠加向量, 待预测商品评分是用户偏好与待预测商品的内积.
- **STAMP**^[18]. 该方法使用商品交互序列全局偏好和局部偏好的叠加向量作为用户偏好, 全局偏好是商品交互序列中所有商品的注意力叠加向量, 局部偏好是商品交互序列的最新交互商品, 待预测商品评分是用户偏好与待预测商品的内积.
- **SR-GNN**^[24]. 该方法使用商品交互序列全局偏好和局部偏好的叠加向量作为用户偏好, 商品特征由 GGNN 生成. 全局偏好是商品交互序列中所有商品的注意力叠加向量, 局部偏好是商品交互序列的最新交互商品, 待预测商品评分是用户偏好与待预测商品的内积.

4.3 评价准则

会话推荐常使用召回率 (Recall) 和平均倒数排名 (MRR) 作为评价准则评价推荐效果. 因而, 本文使用了这两个指标评价模型的性能表现. Recall 评价推荐商品是否命中, 若商品交互序列对应的下个商品出现在推荐商品中, 则为命中, 否则为未命中. MRR 评价推荐商品在命中条件下的排名情况, 若商品交互序列对应的下个商品出现在推荐商品中, 则计算该商品在推荐列表中的排名倒数, 否则记为 0. $@K$ 表示生成推荐列表的长度. 参考领域相关工作的评价方法^[14, 17], 本文采用 Recall@20 和 MRR@20 来评价方法的有效性. Recall@20 和 MRR@20 的计算公式如下:

$$\text{Recall@20} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \text{hit}_i, \quad (14)$$

$$\text{MRR@20} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}, \quad (15)$$

其中, $|Q|$ 表示测试集的数量. hit_i 表示第 i 条测试在长为 20 推荐列表中的命中情况, 命中为 1, 否则为 0. rank_i 表示第 i 条测试在长为 20 的推荐列表中命中下的排名, 若未命中, 则不予计算.

4.4 参数设置

本文使用 python 和 pytorch 共同构建模型. 所使用的显卡是 TeslaV100, 32G 显存. 本文商品隐特征向量的维度是 150, 卷积层的层数为 [1, 2, 3], 学习率为 [0.01, 0.001], 正则化项系数为 $[1e^{-6}, 5e^{-6}, 1e^{-5}]$, 非线性激活函数为 Sigmoid, 训练批大小为 1024.

4.5 比较分析

本小节将在 3 个数据集上进行实验验证, 将所提出方法与相关方法进行比较分析. 表 3 列出了所有方法在 Recall@20 和 MRR@20 上的性能表现. 其中, 比较方法 POP, S-POP, Item-KNN, BPR-MF, GRU4REC, NARM, STAMP 和 SR-GNN 在 Yoochoose 和 Diginetica 两个数据集上的结果来自 SR-GNN^[24], 而在 Globo 数据集上的结果是我们运行得到. 从结果可以看出, IHGCN 方法取得了较好的

表 3 本文所提出方法 IHGCN 与比较方法在 3 个数据集上的结果
Table 3 Results of the proposed IHGCN and the compared methods on three datasets

Method	Yoochoose		Diginetica		Globo	
	Recall@20	MRR@20	Recall@20	MRR@20	Recall@20	MRR@20
POP	6.71	1.65	0.89	0.20	2.27	0.55
S-POP	30.44	18.35	21.06	13.68	2.19	0.38
Item-KNN	51.60	21.81	35.75	11.57	17.28	3.23
BPR-MF	31.31	12.08	5.24	1.98	6.88	1.80
GRU4REC	60.64	22.89	29.45	8.33	30.24	8.97
NARM	68.32	28.63	49.70	16.17	36.70	13.86
STAMP	68.74	29.67	45.64	14.32	38.35	13.95
SR-GNN	70.57	30.94	50.73	17.59	45.62	15.64
IHGCN	71.62	30.56	53.35	18.62	49.38	16.49

性能表现. 在评价准则 Recall@20 上, IHGCN 在 3 个数据集上较于最优的基准方法分别提升 1.5%, 5.2% 和 8.2%; 在评价准则 MRR@20 上, IHGCN 方法在 Diginetica 与 Globo 两个数据集上分别提升了 5.9% 和 5.4%, 仅在 Yoochoose 数据集上的表现稍有不足. 此外, 从传统方法与深度网络方法的角度分析, 实验结果还显示了以下现象.

传统方法. 在所有传统方法中, Item-KNN 方法取得了最好的性能表现, 而 POP 方法则取得了最差的结果. POP 和 S-POP 方法仅仅根据商品在数据集中出现的频次进行推荐, 没有对每个商品交互序列的用户偏好进行个性化分析, 无法捕获用户下一次的行为意图, 因而推荐结果差强人意. Item-KNN 方法使用商品交互序列的最新交互商品作为用户偏好. 此用户偏好实际上为会话的局部偏好, 局部偏好可以反应用户的部分行为意图. FPMC 方法同样把会话的局部偏好作为用户偏好, 因而取得了不错的性能表现. BPR-MF 方法使用商品交互序列所有商品的均值向量作为用户偏好, 没有考虑到每个商品重要性的差异和局部偏好. 本文所提出方法融合了用户的全局偏好和局部偏好, 能够很好地捕获用户的行为意图, 因此取得了优越的推荐性能.

深度网络方法. 由于具有强大的特征抽取能力, 深度神经网络方法相比于传统方法有较大的性能提升. 所提出的 IHGCN 方法和对比的深度推荐方法基本上使用统一的模型框架. 框架由两个模块构成, 分别是商品特征学习模块和用户偏好挖掘模块, 并使基于交叉熵函数和 Adam 优化器优化模型参数. 这一类学习框架可以很好地从短暂的会话中挖掘出用户偏好, 取得较好的推荐性能. 因此, 从表 3 中可以看出, 这些方法取得了比传统推荐方法优越的推荐结果. 此外, 与对比的深度推荐方法相比, 所提出的 IHGCN 方法整体上取得了更好的结果. 这主要是 IHGCN 在商品特征学习模块和用户偏好挖掘模块中均提出了新颖的做法. 在商品特征学习模块中, IHGCN 方法根据商品交互历史与商品间的协同关系构建全局的商品节点图, 并使用图卷积网络从中挖掘出商品的特征表示. 而 SR-GNN 方法根据会话中商品间的序列关系构建局部的商品节点图, 并使用门控图神经网络生成商品特征, 其他方法忽视了商品节点的结构. 因此, IHGCN 和 SR-GNN 方法优于其他深度推荐方法. 在用户偏好挖掘模块中, 除了 GRU4REC 方法, 所有的方法均使用全局偏好和局部偏好的叠加向量. 在挖掘全局偏好时, NARM 方法使用了经 RNN 处理得到的隐藏状态, STAMP 和 SR-GNN 方法使用了粗粒度注意力机制, 而 IHGCN 方法使用特征维度级的细粒度注意力机制. 因此, IHGCN 更能准确地挖掘出用户的全局偏好. 在生成局部偏好时, NARM 方法使用了粗粒度注意力机制, 而 STAMP, SR-GNN 和 IHGCN

表 4 图卷积层数分析

Table 4 Results with different numbers of graph convolution layers

Number of layers	Yoochoose		Diginetica		Globo	
	Recall@20	MRR@20	Recall@20	MRR@20	Recall@20	MRR@20
1	71.58	30.45	52.87	18.31	48.66	15.59
2	71.62	30.56	53.35	18.62	49.38	16.49
3	71.54	30.41	53.04	18.32	47.66	15.67

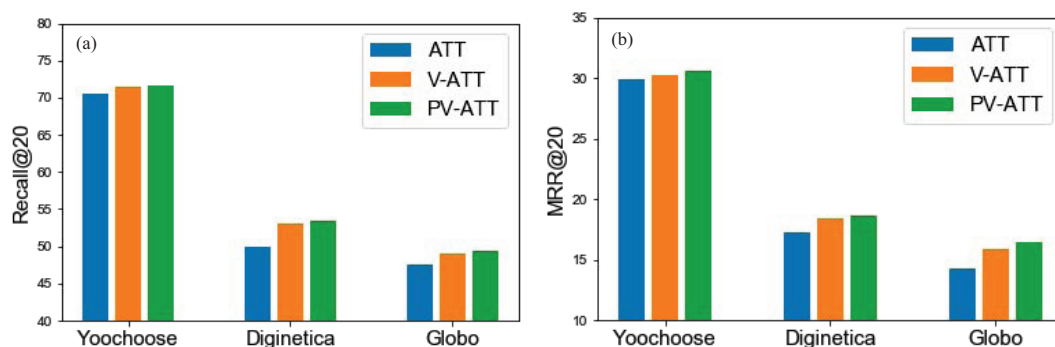


图 3 (网络版彩图) 不同注意力机制的实现效果对比

Figure 3 (Color online) The compared results of different attention schemes. (a) Recall@20; (b) MRR@20

3 种方法仅使用了会话的最新交互商品, 更能体现用户当前的意图. 整体来看, 商品特征使用商品节点的图结构, 全局偏好使用商品交互序列所有商品的权重叠加, 和局部偏好使用会话的最新交互商品, 这一过程可以更好地抓住用户偏好. 综合考虑 3 个数据集上的实验结果, IHGCN 方法取得了较好的性能表现.

综合而言, 本文所提出的方法 IHGCN 通过综合分析商品历史交互信息、用户的全局偏好和局部偏好, 取得了整体最优的性能.

4.6 图卷积层数分析

深度网络模型能够抽取具有强表征能力的特征, 而其网络层数对性能具有较大的影响. 因此, 本节进行了相应的实验分析, 以验证图卷积网络层数对所提出方法 IHGCN 的影响. 表 4 展示了相关结果, 其中图卷积层数设置为 [1, 2, 3]. 整体上, 随着图卷积网络层数的增加, 性能表现随之变好, 继续增加, 性能表现随之下降. 在一定范围内, 增加图卷积层数可以叠加高阶邻居商品节点的特征, 这有利于提取更为准确的商品特征. 然而, 继续增加图卷积层数, 由于更高阶的邻居商品节点与当前商品节点的相似性较低, 这样就容易融入过多的噪声信息, 导致提取的商品特征的准确性降低, 从而导致学习得到的用户偏好也不够准确. 因此, 在实验中, 我们将图卷积网络的层数设置为 2.

4.7 注意力机制对比分析

为了更好地捕获用户的全局偏好, 本文使用了特征维度级的细粒度注意力机制. 本小节进行实验以验证所使用的注意力机制有效性. 图 3 展示了 3 种注意力机制生成全局偏好方法的性能对比, 其中图卷积层数均设置为 2 层. 这 3 种注意力机制分别是粗粒度注意力机制 (ATT) (参考 SR-GNN^[24]), 忽视位置特性的特征维度级的细粒度注意力机制 (V-ATT) (在 SR-GNN^[24] 的基础上使用向量权重), 以

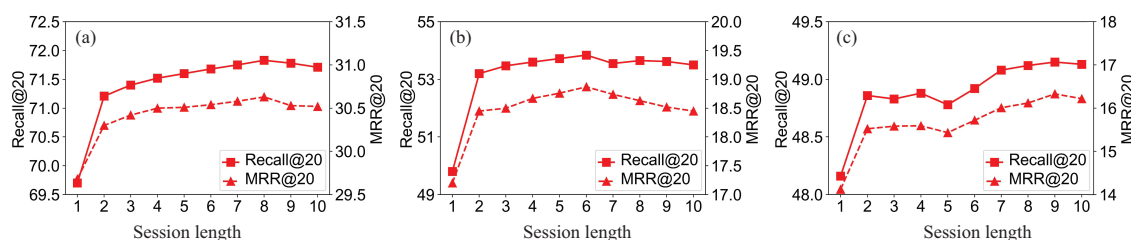


图 4 (网络版彩图) 会话中不同长度序列对推荐结果的影响

Figure 4 (Color online) Results of the proposed method with different session lengths. (a) Yoochoose; (b) Diginetica; (c) Globo

及本文所使用的融合位置特性的特征维度级的细粒度注意力机制 (PV-ATT). 3 种方法均使用 softmax 函数对重要性进行处理.

从在 3 个数据集上的结果可知, PV-ATT 取得了最佳的性能, ATT 的结果最差. 这主要是因为 ATT 采用了粗粒度注意力机制, 在叠加商品交互序列中的商品时使用标量描述商品的重要性, 无法区分不同特征维度的差异. 而 V-ATT 和 PV-ATT 机制使用矢量描述商品重要性, 采用特征维度级别的叠加机制, 能够更好地抓住商品特征分量的贡献差异. 另外一方面, 若商品的点击时间距离最新交互商品的点击时间越远, 该商品的重要性会相应的减弱, 因此商品的位置信息对分析用户的全局偏好尤为重要. 比较 V-ATT 和 PV-ATT 可知, V-ATT 考虑的商品重要性与商品位置无关, 而 PV-ATT 的商品重要性考虑了商品的位置因素. 因此, 本文所提出方法在挖掘用户全局偏好时考虑了商品位置因素, 能够生成更为准确的全局偏好.

4.8 会话长度对比分析

本文所提出方法分析了会话中的商品历史相互信息. 会话长度对推荐结果具有一定的影响. 因而, 本小节进行相关实验讨论会话长度对推荐结果的影响. 推荐结果随着会话长度的变化曲线如图 4 所示. 会话长度在区间 [1, 10] 变化. 从结果可以看出, 推荐结果随着会话长度的增加而逐渐变好. 然而, 当会话长度达到一定值后, 性能变化不大, 有所波动. 这是因为全局偏好使用特征维度级的细粒度注意力机制叠加会话中所有的商品. 随着长度的增加, 全局偏好获得了更多的商品特征, 因而长度的增加有利于抓住用户的喜好. 然而, 继续增加会话序列长度, 用户在会话初期点击过的商品对当前用户偏好的贡献降低, 甚至难以评估用户的当前偏好, 因而性能产生波动或下降.

5 结论

本文提出了一种基于图神经网络的会话推荐模型, 综合分析会话历史交互商品信息, 用户的全局偏好和局部偏好. 根据商品交互历史与商品之间的协同关系构建商品节点图, 并使用图卷积网络提取商品特征. 随后, 使用特征维度级的细粒度注意力机制生成全局偏好, 然后自适应融合全局偏好和局部偏好生成用户偏好. 大量的实验证明了本文方法的有效性. 在未来的工作中, 将进一步探索会话中隐含的商品相关关系, 以形成更优质的商品表示. 此外, 通过结合应用特征, 引入额外的知识, 进一步提升会话推荐的性能.

参考文献

- 1 Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback. 2012. ArXiv:1205.2618
- 2 Du X Y, He X N, Yuan F J, et al. Modeling embedding dimension correlations via convolutional neural collaborative filtering. *ACM Trans Inf Syst*, 2019, 37: 1–22
- 3 He X N, Deng K, Wang X. LightGCN: simplifying and powering graph convolution network for recommendation. 2020. ArXiv:2002.02126
- 4 Cai J P, Lei Y Q, Chen M M, et al. Efficient solution of the SVD recommendation model with implicit feedback. *Sci Sin Inform*, 2020, 50: 1544–1558 [蔡剑平, 雷蕴奇, 陈明明, 等. 带有隐式反馈的 SVD 推荐模型高效求解算法. *中国科学: 信息科学*, 2020, 50: 1544–1558]
- 5 Wu J W, Shen L W, Guo W N, et al. Code recommendation for android development: how does it work and what can be improved? *Sci China Inf Sci*, 2017, 60: 092111
- 6 Qin C, Zhu H S, Zhuang F Z, et al. A survey on knowledge graph-based recommender systems. *Sci Sin Inform*, 2020, 50: 937–956 [秦川, 祝恒书, 庄福振, 等. 基于知识图谱的推荐系统研究综述. *中国科学: 信息科学*, 2020, 50: 937–956]
- 7 Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms. In: *Proceedings of International Conference on World Wide Web*, Hong Kong, 2001. 285–295
- 8 Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry. *Commun ACM*, 1992, 35: 61–70
- 9 Herlocker J L, Konstan J A, Borchers A, et al. An algorithmic framework for performing collaborative filtering. *SIGIR Forum*, 2017, 51: 227–234
- 10 Lin Z Q, Chen H X. Recommendation over time: a probabilistic model of time-aware recommender systems. *Sci China Inf Sci*, 2019, 62: 212105
- 11 Du X Y, Wang X, He X N, et al. How to learn item representation for cold-start multimedia recommendation? In: *Proceedings of ACM Multimedia Conference, Virtual Event*, 2020. 3469–3477
- 12 Andrew A, Chickering D M, Meek C. Using temporal data for making recommendations. 2013. ArXiv:1301.2320
- 13 Rendle S, Freudenthaler C, Schmidt-Thieme L. Factorizing personalized markov chains for next-basket recommendation. In: *Proceedings of International Conference on World Wide Web*, Raleigh, 2010. 811–820
- 14 Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks. 2015. ArXiv:1511.06939
- 15 Cho K, Merriënboer B V, Bahdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches. 2014. ArXiv:1409.1259
- 16 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*, 1997, 9: 1735–1780
- 17 Li J, Ren P J, Chen Z M, et al. Neural attentive session-based recommendation. In: *Proceedings of ACM Conference on Information and Knowledge Management*, Singapore, 2017. 1419–1428
- 18 Liu Q, Zeng Y F, Mokhosi R, et al. STAMP: short-term attention/memory priority model for session-based recommendation. In: *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 2018. 1831–1839
- 19 Quadrana M, Karatzoglou A, Hidasi B, et al. Personalizing session-based recommendations with hierarchical recurrent neural networks. In: *Proceedings of ACM Conference on Recommender Systems*, Como, 2017. 130–137
- 20 Ruocco M, Skrede O S L, Langseth H. Inter-session modeling for session-based recommendation. In: *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, Como, 2017. 24–31
- 21 Jannach D, Ludewig M. When recurrent neural networks meet the neighborhood for session-based recommendation. In: *Proceedings of ACM Conference on Recommender Systems*, Como, 2017. 306–310
- 22 Zhou J, Cui G Q, Zhang Z Y, et al. Graph neural networks: a review of methods and applications. 2018. ArXiv:1812.08434
- 23 Wang X, He X N, Wang M, et al. Neural graph collaborative filtering. In: *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, 2019. 165–174
- 24 Wu S, Tang Y Y, Zhu Y Q, et al. Session-based recommendation with graph neural networks. In: *Proceedings of AAAI Conference on Artificial Intelligence*, Honolulu, 2019. 346–353
- 25 Weimer M, Karatzoglou A, Le Q V, et al. Cofi rank-maximum margin matrix factorization for collaborative ranking.

- In: Proceedings of Advances in Neural Information Processing Systems, Vancouver, 2008. 1593–1600
- 26 Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*, 2009, 42: 30–37
 - 27 Li Z C, Tang J H, Mei T. Deep collaborative embedding for social image understanding. *IEEE Trans Pattern Anal Mach Intell*, 2019, 41: 2070–2083
 - 28 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems, Lake Tahoe, 2012. 1097–1105
 - 29 Tan Y K, Xu X, Liu Y. Improved recurrent neural networks for session-based recommendations. In: Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, 2016. 17–22
 - 30 Wang H W, Guo M Y. Recurrent memory networks: modeling long short-term user preferences for session-based recommendation. *Sci Sin Inform*, 2020, 50: 1867–1881 [王鸿伟, 过敏意. 刻画长短期用户兴趣的基于会话的推荐系统. *中国科学: 信息科学*, 2020, 50: 1867–1881]
 - 31 Li Y, Tarlow D, Brockschmidt M, et al. Gated graph sequence neural networks. 2015. ArXiv:1511.05493
 - 32 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. 2016. ArXiv:1609.02907
 - 33 Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. 2017. ArXiv:1710.10903
 - 34 Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: Proceedings of Advances in Neural Information Processing Systems, Long Beach, 2017. 1024–1034
 - 35 Kingma D P, Ba J. Adam: a method for stochastic optimization. 2014. ArXiv:1412.6980

Item correlation modeling in interaction sequence for graph convolutional session recommendation

Zhao YAN, Xinguang XIANG & Zechao LI*

School of Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China

* Corresponding author. E-mail: zechao.li@njjust.edu.cn

Abstract Session-based recommendation aims to predict the user’s follow-up behavior based on the user’s short-term behavior in the session and is mostly used in scenarios where the user visits the website anonymously. The key problem of session-based recommendation is to model collaborative relationships between items to improve the recommendation performance. Previous methods primarily focused on discovering the interaction sequence of items to learn correlations between items and only explored the information of adjacent interactive items without taking into account historical interactive information. Actually, if two items share the same historical interaction information, they can be considered to have strong relationships. For example, in a session where the interaction sequences of two items are $I_1-I_2-I_3$ and $I_1-I_2-I_4$, the historical interaction information of item I_3 and item I_4 are both I_1-I_2 . Thus, item I_3 and item I_4 can be deemed as similar. Toward this end, this paper proposes a new session-based recommendation model named IHGCN to jointly explore the adjacent interactive information and the historical interactive information by analyzing the interaction sequence of items in the session. To learn the item feature from the interactive information, the proposed method introduces the graph structure to model correlations between items and then uses graph convolutional networks to abstract the item features from the item correlation graph. The proposed IHGCN method introduces a fine-grained attention mechanism in the feature dimension level to discover the global preference of users and then integrate the local preference of users. Based on the learned item features and user preference features, the final recommendation results are obtained. Experiments conducted on three public datasets show the superiority of the proposed method to the state-of-the-art methods.

Keywords session-based recommendation, graph convolutional network, item interaction similarity, user preference, attention mechanism